Replica-Exchange Method Using the Generalized Effective Potential

Soonmin Jang,[1] Seokmin Shin,[1] and Youngshang Pak[2,*]

[1]*School of Chemistry, Seoul National University, Seoul 151-742, Korea*

[2]*Department of Chemistry, Pusan National University, Pusan 609-735, Korea*

(Received 25 April 2003; published 1 August 2003)

We propose an effective scheme for fast conformational searches by combining the replica exchange method (REM) with the generalized effect potential concept. The present method introduces the "q" value from the effective potential as a coupling parameter. It is found that the new method not only requires a much smaller number of replicas than the conventional REM, but also makes it possible to perform effective conformational sampling of complex systems with correct distributions maintained. The advantage of the present method has been demonstrated with *in vacuo* alanine dipeptide using a molecular dynamics simulation.

DOI: 10.1103/PhysRevLett.91.058305 PACS numbers: 82.20.Wt, 36.20.Ey, 87.15.Cc, 87.15.He

Recently, there has been considerable progress in developing efficient conformational sampling schemes for polymers and proteins. Usually, conventional molecular dynamics (MD) or Monte Carlo (MC) simulation methods for biomolecules cannot produce sufficient conformational sampling for canonical distribution at physiological temperature, since systems tend to be trapped in local energy minima due to the nature of the complex potential energy landscapes. In order to circumvent these sampling problems, various simulation strategies have been proposed by introducing non-Boltzmann weighting factors [1–3]. In particular, the replica exchange method (REM) [4,5] is considered to be one of the most promising and efficient methods to sample conformational states of biomolecules [6–8]. In REM, several independent trajectories, called replicas, are generated at different temperatures, and stochastic exchanges between neighboring trajectories are attempted with predetermined intervals during the simulation. The trajectory exchanges between the replicas allow the system to escape from the local energy minima easily, exploring a broad range of the potential energy surface. This scheme is ideal for parallel computation, since each replica can run independently on a different processor and the communication loads between the processors are minimal due to the occasional exchange attempts. The initial version of the method was proposed for an MC scheme and later Sugita *et al.* [7] developed an MD version. Despite its usefulness, the main disadvantage of REM is that it requires a large number of replicas because the energy distributions between neighboring replicas need to have a considerable overlap to have a reasonable acceptance ratio. This is a major obstacle especially with an all-atom based potential. In the conventional REM, the number of replicas scales as $f^{1/2}$, where f is the number of the system's degrees of freedom [9,10]. For example, Zhou *et al.* [11] employed a total of 64 replicas to study all-atom based β-hairpin folding of a small 16-residue peptide in the presence of explicit water. Recently, a couple of variants of REM have been reported [9,12]. Fukunishi *et al.* [10] showed that the number of replicas can be reduced by assigning a different system Hamiltonian to each replica. For protein simulations, as one possible choice they introduced a degree of "hydrophobicity" in the Hamiltonians for "coarse-grained" protein models.

In this Letter, we propose a new Hamiltonian REM for all-atom based protein simulations by introducing the generalized effective potential concept [13–15]. The generalized effective potential $U_q(\mathbf{r}, \epsilon)$ can be obtained by a simple nonlinear transformation of any empirical potential energy function:

$$U_q(\mathbf{r}^N, \epsilon) = \frac{q}{\beta(q-1)} \ln\{1 + \beta(q-1)[U(\mathbf{r}^N) + \epsilon]\}, \quad (1)$$

where $U(\mathbf{r}^N)$ is the original empirical potential in N-dimensional configurational space $\mathbf{r}^N$, $\beta = 1/k_B T$ with Boltzmann constant k_B and temperature T, q is a real number, and ϵ is an adjustable energy shift parameter. The transformed energy reproduces the original potential energy, as the q parameter approaches 1.0. As q is larger than 1.0, this transformation effectively reduces the potential energy barriers, providing much smoother potential energy functions. This has been illustrated in Fig. 1 for a random one-dimensional potential energy [16]. In general, for a given set of q and ϵ, the transformation in Eq. (1) is related to increasing ($q > 1.0$) or decreasing ($q < 1.0$) an effective temperature of the system β', such that [15]

$$\frac{1}{\beta'} = \frac{1}{\beta} + (q-1)\epsilon. \quad (2)$$

In our proposal REM scheme, each replica samples the conformational space on $U_q(\mathbf{r}^N, \epsilon)$, and the corresponding effective Hamiltonian is

$$H_q(\mathbf{p}^N, \mathbf{r}^N, \epsilon) = T(\mathbf{p}^N) + U_q(\mathbf{r}^N, \epsilon). \quad (3)$$

Unlike the ordinary REM where each replica runs independently at different temperatures, in this new REM each replica runs at a single target temperature on

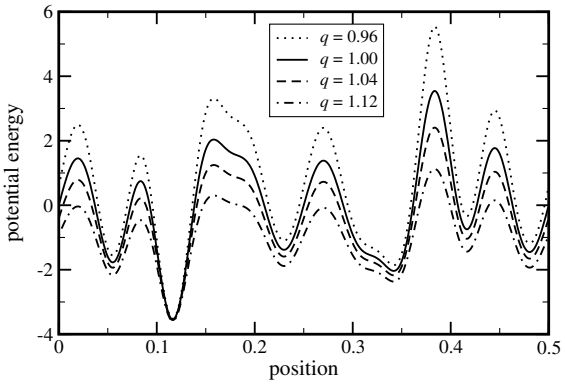

FIG. 1. The random one-dimensional potential energy $U(x)$ and its effective potential $U_q(x, \epsilon)$ defined in Eq. (1) with a choice of $\epsilon = 6.0$ and $k_B T = 1.0$. The solid line is the original potential energy ($q = 1.0$). One can observe that the potential energy barriers are reduced, while the positions of local minima and maxima remain intact, as q becomes larger than 1.0. Also note that the potential energy barriers are more pronounced for $q < 1.0$.

the potential energy surfaces U_q with different q values. Of course, during the replica exchange procedure, the detailed balance condition needs to be satisfied. Following the treatment of Sugita *et al.* [7], the weighting factor of state X consisting of M replicas is given by

$$W_{\mathrm{REM}}(X) = \exp\left[-\sum_{i=1}^{M} \beta H_{q_i}(\mathbf{p}^N, \mathbf{r}^N, \epsilon) \right], \qquad (4)$$

where q_i is the proper q value assigned to the ith replica. Now consider a trajectory exchange procedure between replicas i and j and denote the new state resulting from the exchange attempt as X'. In order to satisfy the detailed balance condition, the transition probability p for the replica exchange needs to satisfy

$$W_{\mathrm{REM}}(X) p(X \to X') = W_{\mathrm{REM}}(X') p(X' \to X). \qquad (5)$$

Thus, the replica exchange probability $p(X \to X')$ is simply given by

$$p(X \to X') = \min[1, \exp(-\Delta)], \qquad (6)$$

where $\Delta \equiv \beta[U_{q_i}(X'_j) + U_{q_j}(X'_i) - U_{q_i}(X_i) - U_{q_j}(X_j)]$ and $U_{q_i}(X_j)$ is the transformed potential energy of the jth replica at state X with q_i.

In our method, the q value for one of the replicas is always set to 1, so that the corresponding replica retains a canonical distribution. Hereby, we denote the ordinary REM as t-REM and the present method as q-REM.

We tested q-REM with a one-dimensional asymmetric double-well model system [17]

$$U(x) = [(x + 1)^2 - 1][(x - 1)^2 - 0.9]. \qquad (7)$$

The analytic expression for the corresponding canonical position probability distribution P is $P(x) = C\exp[-\beta U(x)]$ with normalization factor C. Throughout this study, we have used MD instead of MC as a sampling method and this can be easily achieved by introducing a force scaling factor in the MD scheme [18]. We have used a single particle of unit mass at $\beta = 5.0$. The MD time step was 0.002 in reduced unit and the initial 1×10^5 MD steps were discarded from the sampling to ensure *pseudoequilibration*. A total of 5×10^6 MD steps were included in our MD sampling. The Nosé-Hoover thermostat with a chain length of two [19–21] was used to keep the system at a constant temperature.

Figure 2 shows the position probability distributions obtained from the present scheme, the analytical expression, and the usual canonical MD simulation along with the double-well potential energy given by Eq. (7). With a choice of $\epsilon = 250$, only two replicas were included in our calculation with corresponding q values of 1.000 and 1.005. The replica exchange was attempted every 100 MD steps and the resulting acceptance probability was about 29%. Obviously, the conventional MD trajectory is trapped in one of the potential wells, producing an incorrect probability distribution, while q-REM generates a distribution identical to the analytic one.

Using the all-atom AMBER parm99 force field [22], we applied q-REM to *in vacuo* alanine dipeptide terminally patched with acetyl and N-methyl groups [23–25] at 100.0 K. In general, the empirical potential energy function is given by

$$U(\mathbf{r}^N) = U_{\mathrm{bond}}(\mathbf{r}^N) + U_{\mathrm{angle}}(\mathbf{r}^N) + U_{\mathrm{dihedral}}(\mathbf{r}^N) + U_{\mathrm{nbond}}(\mathbf{r}^N), \qquad (8)$$

where U_{bond}, U_{angle}, U_{dihedral}, and U_{nbond} are the bond stretching, the angle bending, the dihedral rotation, and

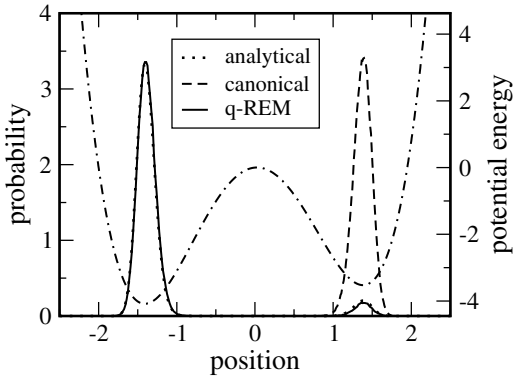

FIG. 2. The model potential of a one-dimensional asymmetrical double well and its position probability distributions from various simulations. The dot-dashed line represents the model potential. The position probability distributions at $\beta = 5.0$ are shown with a dashed line (canonical MD), a solid line (q-REM), and a dotted line (analytic expression). The q-REM distribution is almost indistinguishable from the analytical result.

the nonbond energy terms, respectively. Note that $U_{\text{nbond}}(\mathbf{r}^N)$ consists of both van der Waals and electrostatic interaction terms. Since the dihedral and the nonbond energy are mainly responsible for conformational change, we applied the transformation in Eq. (1) to only the terms U_{dihedral} and U_{nbond}. The resulting effective potential $U_q(\mathbf{r}^N, \epsilon)$ [18] is

$$U_q(\mathbf{r}^N, \epsilon) = U_{\text{bond}}(\mathbf{r}^N) + U_{\text{angle}}(\mathbf{r}^N) + \frac{q}{\beta(q-1)} \ln\{1 + \beta(q-1)[U_{\text{dihedral}}(\mathbf{r}^N) + U_{\text{nbond}}(\mathbf{r}^N) + \epsilon]\}. \quad (9)$$

We performed the simulations starting from a fully stretched conformation. In t-REM, at least a total of five replicas were needed with corresponding temperatures of 100.0, 123.0, 148.0, 178.0, and 213.0 K, respectively. We employed the Berendsen thermostat [26] with a coupling parameter of 0.5 ps. Starting from the structure described above, each replica was subject to another 100 ps MD run with its own temperature without replica exchanges. Then, as a production run, the t-REM simulations were carried out for 4.1 ns with the replica exchanges. After the initial 2.1 ns trajectories were discarded, the data collection was made with a sampling interval of 20 fs (1×10^5 data points). As for q-REM, short trial runs are needed in order to choose an appropriate value for ϵ (for more details, refer to Ref. [18]). Once the ϵ value has been determined, a suitable set of q values can be obtained by trial and error, until a reasonable acceptance ratio (10%–20%) is achieved in the replica exchange. We tested q-REM using two and four replicas with $\epsilon = 30.0$ kcal/mol at $T = 100$ K, and the resulting energy distribution corresponding to $q = 1.0$ is exactly the same for both cases. Thus, in terms of performance, the q-REM with two replicas are comparable to the t-REM with five replicas. We used the q values of 1.000 and 1.002. For both q-REM and t-REM, the replica exchange interval is 0.2 ps and the acceptance ratio is around 10%. Figure 3 shows the energy distributions from q-REM and t-REM. In Fig. 4, we have compared the energy distributions of q-REM at $q = 1.0$ with those of t-REM and conventional MD simulations at 100 K. As one might expect, both q-REM and t-REM results are virtually identical, while the ordinary MD simulation gives a severely biased energy distribution, indicating that the system is trapped at one of the local energy minima. This observation is also confirmed by the conformation distributions of the alanine dipeptide as given in Fig. 5.

The result of the alanine dipeptide system shows that the current method reduces the number of replicas by a factor of 2.5 compared to the conventional REM. As our preliminary study indicates, however, the reduction should be more pronounced with complex systems especially in solvated environments.

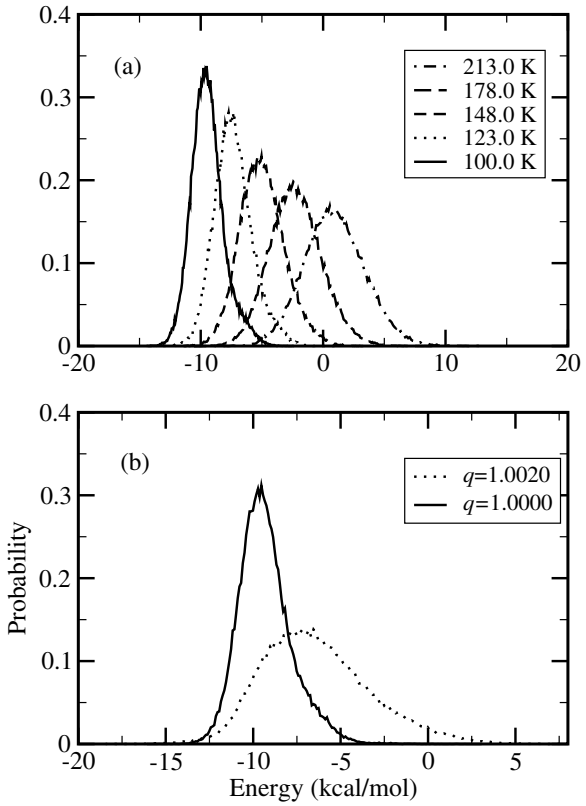

FIG. 3. The energy distributions of alanine dipeptide. (a) t-REM with five replicas and (b) q-REM at 100.0 K with two replicas using the ϵ value of 25.0 kcal/mol.

As was mentioned, the characteristic feature of the transformation in Eq. (1) is to increase/decrease the effective temperature of the system. For $q > 1.0$ ($q < 1.0$), the selective transformation in Eq. (9) represents a targeted "local heating (cooling)" to enhance (suppress) major dynamic motions responsible for conformational changes. However, t-REM relies on global heating (or cooling). As a result, a significant portion of thermal energy is used to excite all the vibrational motions which

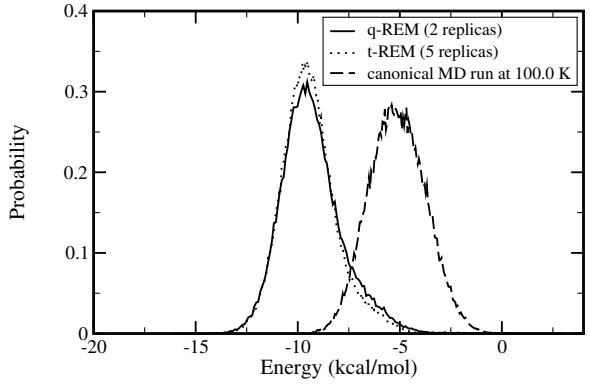

FIG. 4. The energy distributions of q-REM ($q = 1.00$), t-REM, and the normal canonical MD simulations with alanine dipeptide at 100.0 K.

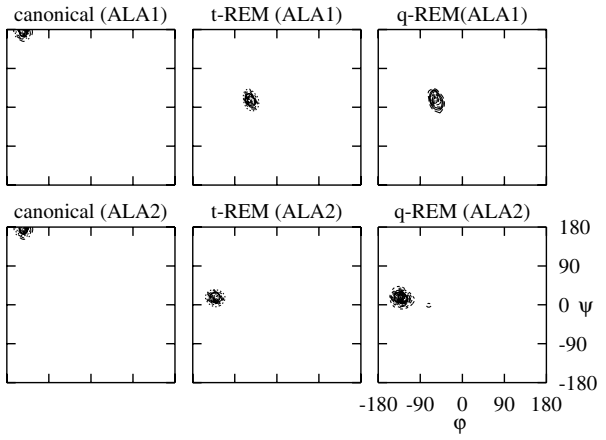

FIG. 5. Ramachandran plot of alanine dipeptide at 100.0 K *in vacuo*. Note that the normal MD trajectory is trapped in an extended conformation, while both *q*-REM and *t*-REM explore almost the same conformational space, which represents the correct distribution of the dipeptide at the simulation temperature.

are not necessarily crucial for conformational fluctuations. Therefore, a broad energy spectrum spanned by REM can be created more effectively by the current strategy. We expect that this new scheme could provide an efficient conformational search tool with considerably fewer CPUs, making it feasible to investigate rather large protein folding problems with all-atom based potentials. Furthermore, one can extend this method by combining it with *t*-REM. Presently, applications of the current method to more realistic protein folding problems are under way.

This work was supported by Korea Research Foundation Grant No. KRF-2002-070-C00048. Y. P. acknowledges the 4th Supercomputing Applications Support Program of Korea Institute of Science and Technology Information. S. J. and S. S. acknowledge support from BK21 Division of Chemistry and Molecular Engineering.

*Electronic address: ypak@pusan.ac.kr

[1] B. J. Berne and J. E. Straub, Curr. Opin. Struct. Biol. **7**, 181 (1997).
[2] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).
[3] Y. Iba, Int. J. Mod. Phys. C **12**, 623 (2001).
[4] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[5] D. Dront, A. Kolinski, and J. Skolnick, J. Chem. Phys. **113**, 5065 (2000).
[6] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[7] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[8] K. Y. Sanbonmatsu and A. E. Garcia, Proteins **46**, 225 (2002).
[9] Y. M. Rhee and V. S. Pande, Biophys. J. **84**, 775 (2003).
[10] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).
[11] R. Zhou, B. J. Berne, and R. Germain, Proc. Nat. Acad. Sci. U.S.A. **98**, 14 931 (2001).
[12] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).
[13] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[14] I. Andricioaei and J. E. Straub, Phys. Rev. E **53**, R3055 (1996).
[15] I. Andricioaei and J. E. Straub, J. Chem. Phys. **107**, 9117 (1997).
[16] R. Zhou and B. J. Berne, J. Chem. Phys. **107**, 9185 (1997).
[17] N. Nakajima, H. Nakamura, and A. Kidera, J. Phys. Chem. B **101**, 817 (1997).
[18] Y. Pak, S. Jang, and S. Shin, J. Chem. Phys. **116**, 6831 (2002).
[19] S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[20] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[21] S. Jang and G. A. Voth, J. Chem. Phys. **107**, 9514 (1997).
[22] D. A. Case, D. A. Pearlman, J.W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman, AMBER 7, University of California, San Francisco, 2002.
[23] S. Ono, N. Nakajima, J. Higo, and H. Nakamura, J. Comput. Chem. **21**, 748 (2000).
[24] T. Terada, Y. Matsuo, and A. Kidera, J. Chem. Phys. **118**, 4306 (2003).
[25] N. Nakajima, Chem. Phys. Lett. **288**, 319 (1998).
[26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).